Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

The Effect of English Learner Reclassification on Student Achievement and Noncognitive Outcomes

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Chin, Mark	Vanderbilt University	0000-0002-0136-2068

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 25 Dec 2020

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Journal of Research on Educational Effectiveness, Volume 14, 2021, Issue 1

DOI or URL to published work (if available) https://doi.org/10.1080/19345747.2020.1831116

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scie through **[Grant number]** R305B1500 to **Institution]** Harvard Unive.The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

The effect of English learner reclassification on student achievement and noncognitive outcomes

Mark J. Chin

Harvard University

Author Note

Correspondence concerning this article should be addressed to Mark J. Chin, Center for Education Policy Research, 50 Church Street 4th Floor, Cambridge, MA 02138. E-mail: mark_chin@g.harvard.edu

Acknowledgements

This work would not have been possible without Glenda Harrell, Matthew Lenard, and Bradley McMillen. I thank Desmond Ang, David Deming, Thomas Kane, Martin West, and seminar participants at Harvard, the Wake County Public School System, and AERA for valuable feedback. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150010 for the Partnering in Education Research Fellowship in collaboration with the Center for Education Policy Research at Harvard University. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education. All errors are my own.

Abstract

English learners' (ELs) day-to-day experiences in school change when reclassified as fully English proficient. Prior research, however, is mixed on how reclassification influences outcomes. Many studies also do not or cannot explore key long-term outcomes or identify impacts over time. In this study I leverage longitudinal student data in a regression discontinuity and find that reclassification after third grade affects ELs' achievement in the short and longer term. Reclassified ELs score considerably higher on mathematics and reading standardized tests in fifth and eighth grade. I also provide the first causal evidence for the impact of reclassification on several theoretically affected noncognitive outcomes. I find that reclassification substantially lowers the level of challenge for work assigned by teachers and increases ELs' out-of-school engagement in the short term. However, effects on noncognitive outcomes attenuate or reverse direction in the longer term. Together, these findings highlight the need for evaluations to consider multiple measures and to identify impacts over time when possible, especially when data on long-term outcomes such as high school graduation, college persistence, or labor market success are unavailable.

Keywords: English learner, regression discontinuity design, policy evaluation, achievement, noncognitive outcomes

The effect of English learner reclassification on student achievement and noncognitive outcomes

More than five million students in US public schools—approximately nine percent of all enrollments—are classified as English learners (ELs) (Musu-Gillette et al., 2017). Because ELs have unique instructional needs, different federal and state laws regulate key features of their education (U.S. Department of Education, 2012, 2016). For example, educational agencies maintain policies that outline how to identify students as ELs, how to support them towards English proficiency, and when to reclassify them as fully English proficient. With the projected growth of ELs in US classrooms, evidence on the impact of these policies is crucial.

Impact evaluations of EL policies largely fall into two broad categories. The first focuses on the various forms of instruction ELs can receive (e.g., bilingual education versus English immersion programs), and the subsequent effect of instruction on outcomes such as test scores and speed to English proficiency (for a review, see Chin, 2015). The second focuses on the effects of being classified as an EL by comparing students who are initially classified upon enrollment to those who are never classified (Shin, 2018; Umansky, 2016), or by comparing ELs who are reclassified as fully English proficient to ELs who are not (e.g., Carlson & Knowles, 2016; Pope, 2016; Robinson, 2011). Results from classification studies in particular provide vital information to policymakers on what happens to students who do not receive the instructional supports associated with being classified as an EL when they might need them.

Observational studies find persistent gaps in educational outcomes between ELs and students not classified as ELs (see Chin, 2015; U.S. Department of Education, 2016).

Naïve comparisons between these two populations, however, conflate the classification effects with other observed or unobserved student characteristics. Recent studies that leverage quasi-experimental designs have provided more rigorous evidence on the impacts of different state and district policies that dictate classification status. Specifically, with several educational agencies determining EL classification using student performance on standardized tests as a key factor, researchers have employed regression discontinuity (RD) designs to credibly identify the unbiased effects of classification on outcomes. This quasi-experimental research has not yielded conclusive evidence, with authors documenting both positive and negative impacts of initial classification on student achievement (Shin, 2018; Umansky, 2016), and positive, negative, and null impacts of reclassification (e.g., Carlson & Knowles, 2016; Johnson, 2019; Pope, 2016; Reyes & Hwang, 2019; Robinson, 2011).

In this study, I build on the literature investigating the impacts of reclassification by answering the following research question: What is the impact of being reclassified as fully English proficient on students' achievement (e.g., standardized test performance) and noncognitive outcomes?

To answer this question, I use administrative data from 3rd-grade EL students in the Wake County Public School System (WCPSS) in North Carolina, one of the largest school districts in the US. In WCPSS (and all of North Carolina), EL students are reclassified if their performance on the annual Assessing Comprehension and Communication in English State-to-State test (ACCESS) assessment indicates English proficiency. Because this policy stipulates that only ACCESS performance dictates reclassification status, I arrive at credibly unbiased estimates of the impact of being reclassified after third grade on student test scores using an RD design. Furthermore, I am uniquely able to identify impacts on a variety of noncognitive outcomes using data from the WCPSS student survey administered each school year. This survey includes measures that district leaders and other researchers have identified as key predictors of success (i.e., in-school engagement, grit, level of challenge for school work, relationships with peers, and familial support), and outcomes beyond those assessed by standardized tests that some hope to also be influenced by students' educational experiences (i.e., future goals/aspirations and civic engagement).

I find that reclassification after third grade substantially increases student reading achievement in the short-term (i.e., two years later) and in the longer-term (i.e., five years later). I also observe increases in mathematics achievement following reclassification in the short- and longer-term; the effect for the former, however, is more attenuated and insignificant. For noncognitive outcomes, I find that reclassified students as fifth graders report considerably higher levels of out-of-school engagement but also that their teachers assign less challenging work. Using an earlier cohort of ELs for which I can observe longer-term noncognitive outcomes, however, I find that these relationships between reclassification and noncognitive outcomes appear to either "fade out" (i.e., attenuate) or reverse completely (for the case of out-of-school engagement).

My study makes several contributions to the existing literature. First, in identifying short-term reclassification impacts for students' noncognitive outcomes in addition to achievement outcomes, I highlight the importance of considering a wide range of measures in evaluations. As others have noted, RD studies on EL classification can be useful because their results demonstrate to policymakers whether existing guidelines for classification should be adjusted or ELs' instructional settings changed (e.g., Robinson, 2011; Robinson-Cimpian & Thompson, 2016). For example, the positive achievement findings from my study by themselves lend support for reclassification thresholds in WCPSS to be lowered so that more ELs realize the gains associated with reclassification. But the findings for students' reported out-of-school engagement in the short-term highlight that focusing solely on achievement may fail to account for the full benefits of reclassification. Contrastingly, the results for students' reported decreased access to challenging schoolwork may predict lower reclassification thresholds to bring other negative consequences—though these downsides may be outweighed by the benefits.

Many RD reclassification studies focus on standardized test performance shortly after reclassification. Though achievement outcomes are crucial measures of student success, they often just serve as proxies for outcomes that educational stakeholders care about but have more difficulty measuring, such as college persistence and labor-market outcomes. Notably, these "ultimate" outcomes are also influenced by a wide range of student noncognitive skills (e.g., Chetty et al., 2011; Heckman, Stixrud, & Urzua, 2006), which are currently understudied in this literature. By examining several measures in this investigation of reclassification's effects, I demonstrate that evaluations of educational policies—related to EL classification or otherwise—are incomplete when considering only certain types of intermediate outcomes, and that this limitation might matter for informing policy when ultimate outcomes are not readily available. Relatedly, by showing that impacts on noncognitive outcomes for students appear to fade out and even reverse as reclassified students progress through school, I highlight the need to evaluate policies' effects over time when possible. If college persistence and labor-market success cannot be measured for students, more proximate measures to these ultimate outcomes (i.e., those collected later in students' careers) may be more ideal to analyze in policy evaluations than those collected shortly after reclassification.

Through use of the unique data from WCPSS' student survey, my study also provides the first RD evidence on what others have highlighted as potential unintended consequences of being classified as an EL. Because of the negative stereotypes associated with classification, researchers note that ELs may experience bias and discrimination when interacting with teachers and peers, and that these experiences may inhibit student success (e.g., Umansky, 2016, 2018). I find that reclassified students report more short-term negative discriminatory interactions with teachers (i.e., lower levels of challenge in the work assigned by teachers) and worse relationships with their peers; for the latter, point estimates are consistently large and negative across models, but less consistently significant. At the same time, reclassified students score higher on an out-of-school engagement composite, which captures students' belief of the importance of school for their futures. In the Background and Conclusion sections, I discuss in more detail how these findings together accord with prior work investigating the unintended consequences of EL classification.

Finally, most existing RD classification studies utilize data from districts in the same state—California (Johnson, 2019; Pope, 2016; Reyes & Hwang, 2019; Robinson, 2011; Robinson-Cimpian & Thompson, 2016; Shin, 2018; Umansky, 2016). As such, more evidence on this topic is needed from contexts with different student populations, classification policies, and instructional supports and settings available to ELs (see also, Carlson & Knowles, 2016). My study is the first to assess the impacts of reclassification policies in North Carolina using data from the state's largest school district—and one of the largest school districts in the country.

In what follows, I review the literature on the effects of reclassification as fully English proficient and the details of the WCPSS context. I then describe the data and methodology I use to identify causal impacts of reclassification. I conclude by sharing and discussing the results, and then consider the policy implications of these findings.

Background

Theoretical Motivation for Investigating Reclassification's Impact

As Robinson (2011) notes, programs serving ELs treat reclassification as one desired outcome: resources are directed to support students towards English fluency and reclassification as fully English proficient. However, for many reasons, reclassification itself should also be seen as an intervention whose impact is worth evaluating. In nearly all contexts, being reclassified changes the day-to-day experiences of students in schools. These changes may be intended (e.g., losing access to English Language Development instruction and linguistically accessible core academic content) or unintended (e.g., removal of the EL label and any associated stigma; reintegration with English-speaking peers in mainstream classrooms; Umansky, 2018), which can subsequently affect reclassified students' outcomes. The direction and magnitude of reclassification's impact will likely vary—even for the same student over the course of his or her career—due to differences in policies and the supports available to near- or just-English proficient ELs across contexts (e.g., different districts; middle school versus elementary school). Notably, providing the programming necessary for ELs to become fluent in English comes at financial cost. Assessing the impact of reclassification may provide decision-makers guidance on whether more (or fewer) near-proficient ELs should be reclassified—with implications for how the resources that would be devoted to their specialized instruction could instead be diverted to ELs farther from English fluency.

Existing Empirical Studies on Reclassification and Student Outcomes

Most recent studies measuring the impact of being reclassified as fully English proficient explore ELs' achievement outcomes. In the states and districts where these evaluations have occurred, performance on a standardized test, relative to different scoring thresholds, typically comprises one factor dictating students' reclassification status. As such, the authors of these studies arrived at credibly causal estimates by leveraging these classification policies in RD designs.

Results from reclassification RD studies—which often consider both English language arts (ELA) and mathematics achievement, occur in a variety of contexts, and span several grade levels—are mixed. Robinson (2011), for example, finds null impacts of reclassification on year-after mathematics and ELA standardized test scores for ELs in elementary and middle school, but negative impacts for those reclassified in high school. Contrastingly, in their studies of middle schoolers who are reclassified, Johnson (2019) and Reyes and Hwang (2019) find null impacts on high school test achievement. Pope (2016) finds that those who are reclassified actually experience positive gains on standardized tests, though this result holds true primarily for ELs reclassified in earlier grades and only for ELA achievement. Finally, whereas all these aforementioned studies focus on districts in California, Carlson and Knowles (2016) leverage data from Wisconsin and find that reclassification in tenth grade improves individuals' performance on the reading and English portions of the ACT.

Less frequently, some RD studies have explored the effects of classification policies on more proximal measures to key long-term outcomes (e.g., labor-market success), such as high school and post-secondary outcomes—again with mixed results. Carlson and Knowles (2016) find that reclassification in tenth grade leads to higher probabilities of enrollment in a postsecondary institution, with suggestive positive impacts on high school graduation rates. Johnson (2019) similarly finds some evidence that reclassification in eighth grade increases the chance that students are on track for graduation in high school. Contrastingly, Robinson-Cimpian and Thompson (2016) find a negative effect of reclassification in high school on graduation rates in the Los Angeles Unified School District though, once the standards for reclassification were raised, this effect disappeared. Finally, a small subset of reclassification studies has explored other intermediate predictors of students' future success using RDs. Reyes and Hwang (2019) and Robinson (2011) both find null effects of reclassification on course enrollment patterns and attendance, though Umansky (2018) does find evidence that ELs reclassified in fifth grade are more likely to take a full course load (i.e., enrollment in a sixth grade mathematics, science, and ELA class) the following year.

The Current Study

As many have noted, the impact evaluation literature on EL reclassification policies suggests to educational policymakers that, in considering the merits of their own classification guidelines, context is key. Significant variation exists within the same state (Cimpian, Thompson, & Makowski, 2017) and achievement impact estimates—even from the same district—can be influenced by changes to classification policies (Robinson-Cimpian & Thompson, 2016). Positive, negative, and even null impacts of policies can emerge in any given district as estimates clearly depend on the interplay of the student population under consideration (e.g., demographic composition, grade level), the instructional settings available to ELs, and the policies themselves.

Robinson (2011) argues, however, that of potential RD estimates, precise null effects are the most desirable because non-null effects indicate that a better, alternative policy exists. For example, non-null effects on achievement imply that the level of English fluency of reclassified ELs does not match the level required to succeed academically without instructional supports (Robinson, 2011). Negative effects thus suggest that some who exit EL status should not be reclassified, and positive effects suggest that some ELs should have exited earlier.

Robinson (2011) also stresses, however, that null effects, though better than non-null effects, do not indicate that the existing reclassification policy is the best possible policy. For example, reclassification may demonstrate a null impact on students' test performance the year after losing EL status (see Robinson, 2011), but estimates may change as students progress through school, and reclassification may affect other measures that predict students' long-term success. Figure 1 depicts a simple schema that illustrates why investigating the impacts of reclassification on multiple measures over time is key to determine the appropriateness of any given reclassification policy.

[Insert Figure 1 about here.]

Ideally, decision-makers would know the impact of reclassification on their ELs' ultimate outcomes (i.e., the box in Figure 1 furthest from the time of reclassification). Indeed, some of the aforementioned studies have explored these outcomes (primarily high school graduation), but typically for ELs reclassified in later grades. For most evaluation research, identifying the impact of reclassification on ultimate outcomes is implausible due to lack of access to such data (i.e., college persistence or labor market-success) or because analyzing this data would require waiting several years (i.e., for ELs reclassified in earlier grades). Identifying the impacts of reclassification on key proximal student outcomes can thus help inform policymaking in the short-term.

However, the above-referenced research on the impacts of reclassification have primarily focused on the effects of policies on short-term achievement outcomes. Looking at Figure 1, this focus ignores longer-term achievement outcomes (which are more proximal to the desired long-term outcomes) and noncognitive outcomes in the short and longer term. With several studies highlighting the importance of noncognitive skills for long-term outcomes (Heckman et al., 2006), even when an intervention's impact on achievement outcomes fades out (Chetty et al., 2011), studies would ideally consider reclassification's impact on a range of intermediate measures. If these assessments in totality show null effects, educational agencies should feel more convinced that their policies are appropriate.

As such, in this study I investigate the impacts of reclassification in WCPSS for 3rd-grade ELs on a broad array of noncognitive outcomes, in addition to considering the student achievement measures that many prior studies have focused on. Without immediate access to long-term outcomes for most of these younger students, even outcomes typically available to the district (e.g., graduation rates), considering a range of intermediate measures is even more integral to this evaluation. The noncognitive outcomes I consider include those that leaders in WCPSS have identified as key to their students' outcomes, as well as those that prior research has suggested to be influenced by

EL classification and to matter for long-term success. Notably, despite their importance, no existing study has considered them in evaluations.

The first set of noncognitive outcomes include three measures of students' experiences in school. Specifically, I explore how reclassification affects students' reported level of challenge for work assigned by teachers, relationships with peers, and their in-school engagement (e.g., relationships with teachers; perceptions of the relevance of schoolwork). In-school relationships contribute to the establishment of positive learning environments (e.g., Appleton, Christenson, Kim, & Reschly, 2006; Poulou, 2017) and a student's peer networks can have significant influence over their learning outcomes (e.g., Hoxby, 2000).

EL classification, however, may affect the development of such relationships through several pathways, which may subsequently have repercussions for ELs' success. For example, though an intended consequence of classification is to ensure that students who need targeted instruction to develop English proficiency and more easily access academic content receive it (Lau v. Nichols, 1974), targeted instruction might alter ELs' relationships with teachers and peers if schools isolate them from non-EL students and mainstream classrooms in order to provide it (Umansky, 2018). This practice may cause ELs to experience stigma (Dabach, 2014) and/or encourage lower expectations from teachers (Blanchard & Muller, 2015; Thompson, 2015), which itself could lead to less challenging work being assigned to or further tracking into less rigorous courses of these students (Harklau, 1994; Kanno & Kangas, 2014). Alternatively, some argue that this special programming associated with classification allow for ELs to develop deeper relationships and social ties to others in school who share similar experiences, despite potential negative academic consequences (Harklau, 1994; Suárez-Orozoco, Suárez-Orazoco, & Todorova, 2008). Finally, some combination of these pathways may most accurately capture the impacts of classification in settings that isolate ELs: reclassified students may both experience less stigma from peers and teachers *and* weaker in-school connections after moving entirely to mainstream classrooms.

Even in classrooms where ELs learn alongside their non-EL peers, merely being classified may affect peer and teacher expectations and, subsequently, the development of relationships (Link & Phelan, 2001; Umansky, 2016). Again, however, the predicted direction of reclassification's impact is unclear. Teachers of mainstream classes may hold lower expectations for ELs (Dabach, 2014) or they may invest additional effort to ensure these students' success. Indeed, research shows that certain instructional approaches (i.e., bilingual versus English immersion settings) appear to lead to more positive achievement outcomes for ELs (e.g., Valentino & Reardon, 2015) and less negative teacher perceptions, potentially because of the asset-orientation of the teachers in these classrooms (e.g., Umansky & Dumont, 2019; Yoon, 2008). The impact of reclassification on these key in-school experience noncognitive measures is thus an empirical question with, to date, limited causal evidence.

In my evaluation, I also include a second set of noncognitive outcomes—students' grit, out-of-school engagement (e.g., civic engagement; future aspirations and goals), and familial support for two primary reasons. First, studies show that the first two measures predict student achievement and attendance (e.g., Duckworth, Peterson, Matthews, & Kelly, 2007; West et al., 2016), and some argue that schools should prepare students to be engaged citizens (Galston, 2007). Second, the aforementioned effects of labeling, stigma, and lower expectations associated with classification may disengage ELs in school (Dabach, 2014; Reyes & Hwang, 2019; Thompson, 2015). This disengagement could lead students to place a lower value on education for future opportunities or exhibit less persistence and perseverance in the face of adversity (i.e., grit).

I provide more detail on the measurement of each of the six student noncognitive outcomes used in my study below.

Method

District Context

Between the 2008–2009 and 2016–2017 school year, enrollment in WCPSS grew from approximately 140,000 students to about 163,000 students. However, in the same timeframe, the number of ELs remained fairly steady, averaging 10,000 to 13,000 a year; most of these students were enrolled in elementary school.

Every spring, ELs enrolled in the district take the ACCESS exam. Performance on this exam, which is used by 36 different states and the District of Columbia, determines students' English proficiency. In WCPSS and North Carolina, students must meet three different ACCESS thresholds for performance to be successfully reclassified as fully English proficient. ELs must specifically receive proficiency level scores of 4.0 ("expanding" proficiency) in reading ("Students at this level generally can understand written language related to specific topics in school, for example […] distinguish view points and justifications described in editorials and other written texts") and writing ("Students at this level generally can communicate in writing in English using language related to specific topics in school, for example […] narrate stories with details of people, events and situations"), and a score of 4.8 on an overall composite scale (a weighted combination of the four main ACCESS scales, described in more detail below; WIDA, 2019). WCPSS ELs do not need to meet any performance thresholds for the listening or speaking ACCESS scales to be reclassified.

Of approximately 30,000 students that took at least one ACCESS test between 2008–2009 and 2016–2017, 56 percent would have been reclassified based on ACCESS performance alone in at least one test administration.[1] Of students who met each reclassification score threshold, the largest proportion did so in third grade (37.8 percent) or fourth grade (15.8 percent). Very few ELs in the district are reclassified before third grade. I provide additional details on the characteristics of the ACCESS exam below.

In WCPSS, being reclassified as fully English proficient means explicitly the loss of individual language assistance. Depending on the English proficiency of the student, the intensity of this assistance varies across three levels: (from most intense to least) comprehensive, moderate, and transitional. Final authority over the intensity and type of language supports ELs receive rests with WCPSS teachers. However, in recent years the district has provided more guidance on who should be placed at each level, what supports should be provided to students at each level, and encouraged teachers to use other empirical data (i.e., student performance on the initial EL placement test and the ACCESS exam) to inform their judgments. For example, the district recommends that elementary school students—the focal population of this study—requiring comprehensive language support attend a scheduled English as a Second Language class as opposed to receiving in-class support. Transitional ELs, on the other hand, meet less frequently with EL staff and receive less structured language supports; these EL staff instead primarily

[1] As I demonstrate below, compliance to this reclassification policy is high in WCPSS.

work with core school staff to scaffold transitional ELs into mainstream instruction. As noted earlier, the purported loss of these language supports—in conjunction with other changes to reclassified students' in-school experiences (i.e., removal of the EL label and any associated stigma; reintegration with English-speaking peers in mainstream classrooms)—motivates treating reclassification as fully English proficient as an intervention worth evaluating.

Data Overview

To investigate the impact of reclassification as fully English proficient on student outcomes, I use administrative data collected by WCPSS starting in 2008–2009 (the first school year the district used the ACCESS test to determine English proficiency) through 2017–2018. From this data, I extract longitudinal records for all students ever classified as an EL in third grade.

I focus on third grade ELs for several reasons. First, as noted above, the majority of EL students in the district whose ACCESS scores alone suggest English proficiency should be reclassified starting after third grade. Second, in 2014–2015, WCPSS began surveying students in fifth, eighth, and ninth grade to measure their perceptions of their learning experiences; I describe the student survey in more detail below. By focusing on third grade ELs, I can thus identify the shorter- and longer-term impacts of reclassification on surveyed outcomes (in addition to performance on end-of-year [EOY] standardized tests, which students take beginning in third grade) for several cohorts of students in my data. Finally, prior evidence suggests that reclassification during upper elementary grades specifically may have significant consequences. Pope (2016) found positive effects of reclassification that grew with each year, but only for ELs reclassified in second through fourth grade.

Key Data Sources

The ACCESS Exam

Though the district uses ACCESS proficiency level scores from the reading, writing, and overall composite domains of the test to determine reclassification status, ELs are assessed on and receive several scores for other domains.

The ACCESS test itself comprises multiple-choice questions and constructed-response tasks (WIDA, 2019). Performance on these items determine raw scores (i.e., the actual number of items correctly answered) in the primary domains of listening, reading, writing, and speaking. Because raw scores do not account for item difficulty, the group that develops, administers, and scores the ACCESS—the WIDA Consortium—also provides scale scores on these four domains. Scale scores take into consideration item difficulty and are thus comparable across grades and test administrations—making them ideal to evaluate individuals' growth in English proficiency over time. WIDA stresses, however, that scale scores are not comparable across domains (WIDA, 2019).

To help interpret scale scores, WIDA also provides proficiency level scores. These scores, ranging from one through six (with decimal scores in between), map ACCESS performance onto a set of standards to aid in evaluating proficiency, and specifically capture student statuses of "entering", "emerging", "developing", "expanding", "bridging", and "reaching" English proficiency. Proficiency levels are grade specific, preventing true evaluations of growth over time for individuals, but do allow for comparisons across domains.

Finally, ACCESS test-takers receive several composite scale scores—weighted sums of scores from the four primary ACCESS domains. The oral language composite score weights listening and speaking scores equally; the literacy composite score weights reading and writing scores equally; the comprehension composite score weights listening and reading scores at 30% and 70%, respectively; and the overall composite score weights listening, speaking, reading, and writing at 15%, 15%, 35%, and 35%, respectively. These composite scale scores are again converted into interpretable proficiency levels for WIDA partner agencies.

Potentially useful variation in ELs' English proficiency is lost (though interpretability is gained, as well as comparability across scales) when using the ACCESS exam's (coarser) proficiency level scores instead of scaled scores; as such, in my study, I focus on scale scores. Even though WCPSS focuses on proficiency levels, because proficiency levels map directly to non-overlapping scale score ranges, I can easily identify the comparable scale score thresholds that would determine reclassification. This decision, however, introduces another problem: without any adjustments, scale scores are not comparable across domains (WIDA, 2019). The RD approach I employ to identify impacts on outcomes hinges on finding a single score that plausibly determines reclassification.

As such, in order to leverage information from ELs' (non-comparable) scale scores on each domain that WCPSS uses to determine classification status via proficiency levels (i.e., reading, writing, and overall composite), I adjust these scores in order to follow Reardon and Robinson (2012) and use a binding-score RD (see also, Johnson, 2019; Robinson, 2011).

Specifically, I first identify the reading, writing, composite scale scores that map onto the corresponding proficiency level score threshold for reclassification in WCPSS (4.0, 4.0, and 4.8, respectively). I then predict that ELs who score at or above this new threshold to be reclassified as fully English proficient moving forward, and those that score below it to be continuing EL services. Next, I rescale these scale scores for analyses by: 1) subtracting the reading, writing, and composite scale score reclassification threshold from each EL's respective scale scores (i.e., within domain, those scoring at the threshold receive scores of zero), and 2) dividing these re-centered reading, writing, and composite scale scores by each domain's standard deviation, so that differences in scores across domains are on the same scale. Finally, I take the lowest re-centered, standardized scale score across the domains of reading, writing, and composite—the scores determining reclassification—for each EL (henceforth referred to as each EL's "ACCESS" score for simplicity).

In Figure 2, I show that these ACCESS scores predict WCPSS students' actual EL status very well in the following year. Nearly all students scoring above the ACCESS score threshold are actually reclassified, and nearly all students scoring below the threshold are not.

[Insert Figure 2 about here.]

The WCPSS Student Survey

The student survey administered each spring queries students on approximately 50 different items, and their responses to these items can be used by WCPSS administrators in evaluations and planning (Huang, 2018). To simplify analyses, the district categorizes these items into eight domains: teacher-student relationships, rigor scale, control and relevance of schoolwork, peer support for learning, future aspirations and goals, civic engagement, family support and learning, and Duckworth Grit scale.

I conducted exploratory and confirmatory factor analyses (EFA, CFA) using the 5th- and 8th-grade student survey responses from my sample to identify a parsimonious set of noncognitive outcome constructs for analyses. EFA suggested a five-factor solution when extracting only factors with eigenvalues greater than one (Kaiser, 1960). Though individual item loadings were generally acceptable following a promax rotation (i.e., . $\geq$ .35; see Online Appendix Table 2), one survey item weakly loaded onto all five factors: "My teachers give me challenging work" (*challenging work*). As such, in CFA, I did not have this item load onto any noncognitive factor; all other items I had load onto the factor for which they were most strongly related to.

Goodness-of-fit statistics from CFA suggested that the fit of the data-driven five-factor solution for noncognitive outcomes was acceptable (see Online Appendix Table 3; Hu & Bentler, 1999). Furthermore, when comparing fit statistics (i.e., AIC and BIC) between the data-driven five-factor solution and the proposed categorization by WCPSS, the former was shown to fit the data better. As such, to estimate ELs' noncognitive outcomes, I averaged their responses across items within each recommended data-driven composite, and then rescaled students' average scores as *z*-scores within grade and year for use in analyses. These five constructs (excluding the *challenging work* item) capture: *in-school engagement* (e.g., "At my school, teachers care about students"; "Most of what is important to know you can learn in school"), *grit* ("When I do schoolwork, I check to see whether I understand what I am doing"; "I finish whatever I begin"), *out-of-school engagement* ("I believe I can make a difference in my community", "I plan to continue my education following high school”), *peer relationships* (“Other students at school care about me”, “I enjoy talking to the students here”), and *family support* (“My family/guardian(s) are there for me when I need them”). In Table 1 of the Online Appendix, I provide reliabilities and item text for these measures.

Key Variables

Annual student data from WCPSS can be divided into five distinct categories. In my analyses I use student *demographic data*, which include indicators for gender, race, identification as Hispanic, disability status, and English learner classification status. The second category contains information on student performance on the ACCESS test, which is administered annually in the spring. In WCPSS the ACCESS test helps determines whether ELs are reclassified. This *ACCESS test data* include the grade level of the test and students’ ACCESS scale scores on several domains; as mentioned above, I focus on a score for ELs’ performance derived from the reading, writing, and overall composite domains.

The third category of data I use comprises student *achievement outcomes*—student performance on EOY standardized tests. To assess the shorter- and longer-term impacts of reclassification following third grade on student achievement outcomes, I focus specifically on the 5th- and 8th-grade mathematics and reading EOY standardized tests. I rescale students’ achievement outcomes as *z*-scores within grade and year.

Along with students’ demographic data, ACCESS test data, and achievement outcomes, I also consider responses to the 5th- and 8th-grade WCPSS student survey. Above I describe how I derive composites from this survey data that measure different student *noncognitive outcomes* to use in my evaluation. The first set of outcomes describe students' experiences in school, and include the aforementioned measures of challenging work, in-school engagement, and peer relationships. The second set include students' self-reported out-of-school engagement, familial support, and grit.

Analytic Methods

Due to unobservable factors, a simple comparison of outcomes between ELs and those reclassified will likely yield a biased estimate of the effect of reclassification. As such, I use an RD approach to identify this effect, leveraging the WCPSS policy of reclassifying ELs as fully English proficient only after they score above a set of thresholds on the annual ACCESS test. Using the RD approach, I can arrive at plausibly causal estimates of the impact of reclassification for EL students scoring around the ACCESS reclassification thresholds under the assumption that students on either side of the cutoff, who should experience different EL classification outcomes, are similar to one another on all other unobservable and observable characteristics. Below I describe the model for my RD, and present evidence supporting the validity of this assumption.

Regression discontinuity model. To measure the impact of reclassification as fully English proficient on student achievement and noncognitive outcomes, I estimate:

$$Y_i = f(A_{iGrade3}) + \delta C_{iGrade3} + \boldsymbol{X}_i'\boldsymbol{\theta} + \varepsilon_i \qquad (1)$$

In this model, i indexes students. Depending on the model, the outcome variable, Y_i, reflects student performance on the achievement outcomes or noncognitive outcomes described above. $A_{iGrade3}$ captures EL students' 3rd-grade ACCESS score.[2] As such, $f(A_{iGrade3})$ represents some function linking this score to outcomes. I use a local linear regression (but also test quadratic relationships as sensitivity checks) to model this

[2] Because students' ACCESS scores are discrete, in all models controlling for these scores I follow Lee and Card's recommendation (2008) and cluster standard errors at the score level.

relationship and allow the relationship to vary on either side of the reclassification thresholds.

In order to limit the influence of outliers (i.e., ELs who way underperform or way overperform the reclassification threshold) on impact estimates, I restrict the bandwidth of ACCESS scores used in analyses. Specifically, I estimate all RD models using the *rdrobust* command, whose default is to consider only observations within an optimal bandwidth, determined using a completely data-driven process (Calonico, Cattaneo, & Titiunik, 2014).[3] I defer to the optimal bandwidth for each outcome, but also report results for bandwidths of 1.2 and .8 times the size of the optimal value (see Robinson-Cimpian & Thompson, 2016) as sensitivity checks.

$C_{iGrade3}$ indicates whether a student scores above the ACCESS score reclassification threshold in third grade. Students who do should no longer be ELs starting in fourth grade, and those who fail to meet it should. I discuss the coefficient on $C_{iGrade3}$ in more detail below.

Finally, I include a vector of control variables in my RD model to improve the precision of estimates. This vector contains data on student gender, race, identification as Hispanic, and disability status (all collected prior to taking the 3rd-grade ACCESS exam). It also includes dummy variables for: the interaction between the school year each student took the 3rd-grade ACCESS exam and the domain (i.e., reading, writing, or composite) of each EL's lowest scale score; and linear terms for each student's scale scores on the four primary ACCESS domains and the overall composite domain (for similar models, see Johnson, 2019; Robinson, 2011).

[3] I also use a triangular kernel to estimate models using *rdrobust*, also the default, but present results without any weighting (i.e., a uniform kernel).

Regression discontinuity instrumental variables model. If "compliance" between students' ACCESS performance in the third grade and their classification status following the test were perfect, the δ coefficient from equation (1) would capture the impact of reclassification on student outcomes. In practice, however, performance on the 3rd-grade ACCESS does not solely determine students' EL status over time. Specifically, many students whose ACCESS score in third grade places them just below a threshold become reclassified in a later grade—perhaps unsurprisingly, as these ELs initially performed near English proficiency. To observe this pattern, see Figure 3, which plots different measures of students' reclassification status against their 3rd-grade ACCESS score.

[Insert Figure 3 about here.]

In Figure 3, I show that most students who meet the ACCESS reclassification threshold in third grade are reclassified as fully English proficient in subsequent grades. Importantly for the RD design—which hinges on differences in the running variable resulting in differences in treatment receipt—I also observe a sharp discontinuity in reclassification status for students just on either side of the reclassification threshold. However, many students who fail to pass the ACCESS reclassification threshold in third grade do ultimately get reclassified in one time point following the 3rd-grade ACCESS test. In fact, more than half of students performing right below the threshold in third grade lose their EL status in fifth grade, and more than 80 percent achieve reclassification in some grade from fifth through eighth grade. The δ coefficient from the model represented by equation (1) thus specifically captures the impact of scoring above the

3rd-grade reclassification threshold on student outcomes, and should be interpreted as an estimate of the intent-to-treat (ITT) parameter for reclassification.

In order to address this treatment "non-compliance" when estimating the effect of reclassification on student outcomes, I combine the RD approach described above with a two-stage least squares (2SLS) instrumental variables approach (RD-IV). To implement this RD-IV, again estimated using *rdrobust,* I first predict each student's EL status over time following the 3rd-grade ACCESS:

$$RC_i = f(A_{iGrade3}) + \delta C_{iGrade3} + \boldsymbol{X_i'\theta} + \varepsilon_i \tag{2}$$

Depending on the model, RC_i captures either each student's total number of years reclassified between fourth and fifth grade or the number of years reclassified between fourth and eighth grade.[4] I use the same covariates in the model represented by equation (2) as I use in the model represented by equation (1). I then estimate the following model:

$$Y_i = f(A_{iGrade3}) + \lambda\widehat{RC_\iota} + \boldsymbol{X_i'\theta} + \varepsilon_i \tag{3}$$

$\widehat{RC_\iota}$ represents each student's predicted number of years reclassified as fully English proficient in the years after the 3rd-grade ACCESS test. This prediction depends solely on variation caused by whether a student initially just passes the ACCESS reclassification threshold. The λ coefficient thus captures the causal impact of each additional year of reclassification on outcomes for students around and on either side of the reclassification threshold, and who "comply" with the WCPSS' policy for determining reclassification based on ACCESS performance in the third grade.[5]

[4] Because nearly all ELs who scored just below the ACCESS reclassification thresholds in third grade eventually achieved reclassification status by eighth grade, I focus on the years reclassified instrumented variable.

[5] As I describe in my conclusion, though these estimates may be limited in their generalizability, policymakers will find them useful in choosing reforms that influence the educational experiences of ELs.

Of the few reclassification studies described above, only Pope (2016) explored the importance of time reclassified as fully English proficient for achievement. By instrumenting for years reclassified (as opposed to just ever being classified), I thus build on Pope's work and investigate the extent to which it matters that ELs who score nearly identically on the ACCESS in the third grade can differ by up to five years in time reclassified by eighth grade. It is important to stress here, however, that λ captures the *average* impact of each additional year of reclassification between either third and fifth grade or third and eighth grade. Put differently, unless I assume linearity in effects, the effect of each year of reclassification on students' outcomes may vary over time. In certain grades, reclassification may matter more for achievement or noncognitive outcomes than in others.

Internal validity of estimates. To support the internal validity of my RD estimates of reclassification's impact on outcomes, I must adequately address several key challenges (Jacob, Zhu, Somers, & Bloom, 2012). First, I must show that students who pass the ACCESS score threshold demonstrate observable differences in EL status when compared to those who fail to pass. In Figure 3 I provided convincing visual evidence that this was true in WCPSS. ELs whose 3rd-grade ACCESS score was at the respective reclassification threshold attained significantly more years of reclassification than those who scored just below the threshold. I perform a formal empirical test of this below.

Second, I must demonstrate that students' ACCESS scores are independently determined from the reclassification thresholds. If EL students (or some other educational stakeholder) can manipulate this score to ensure achieving (or not achieving) reclassification status, this would challenge the validity of estimates. Figure 4 provides visual evidence suggesting no manipulation.

[Insert Figure 4 about here.]

From Figure 4, I can examine the smoothness of the distribution of 3rd-grade ACCESS scores around the reclassification threshold. If I observe a spike of ELs scoring right at the reclassification threshold (i.e., at zero) or right below the threshold, this might indicate that strategic behavior had occurred in WCPSS to manipulate students' EL status. Visually, this does not appear to be the case. Though a common statistical test for this stacking at the cutoff fails (McCrary, 2008) in my data, this may due to the discreteness of ACCESS scale scores (Fransden, 2017). Qualitatively, score manipulation is unlikely for several reasons. First, personnel outside of WCPSS score the ACCESS tests. Second, the ACCESS scores I use in the RD models leverage students' performance on three different scales (i.e., the reading, writing, and overall composite domains), one of which itself is a weighted sum of scores from the four primary ACCESS domains (i.e., overall composite domain), making intentional score manipulation difficult.

Finally, to support the internal validity of RD estimates, EL students on either side of the ACCESS reclassification threshold would ideally be similar. Though I obviously cannot test for equivalence on unobservable characteristics, I can test for differences on observable characteristics. Evidence of similarity on observable characteristics may assuage concerns that ELs and those just reclassified systematically differ from one another such that RD estimates may be biased by other factors, i.e., by differences in other characteristics and/or treatments occurring at the ACCESS score reclassification threshold that influence students' outcomes (but not through reclassification itself).

I first look for differential inclusion in the final analytic samples for my RD models. To arrive at my final analytic samples, I make a series of restrictions. First, I exclude 3rd-grade EL students who do not have scores for all four of the primary ACCESS domains (i.e., reading, writing, listening, and speaking) in addition to the overall composite, do not take the 3rd-grade ACCESS test, and do not have data on demographic control variables (i.e., disability status, gender, race, and identification as Hispanic) before taking the 3rd-grade ACCESS exam. This decision excludes 1,352 of 12,656 (11 percent) third graders who were ELs between 2008–2009 and 2015–2016.

I then create four subsamples of 3rd-grade ELs, each excluding an additional set of students. These four subsamples comprise students with mathematics and reading EOY test scores in fifth grade (*G5 Achievement Sample*), mathematics and reading EOY test scores in eighth grade (*G8 Achievement Sample*), mathematics and reading EOY test scores in fifth grade and all student survey outcomes in fifth grade (*G5 Survey Sample*), and mathematics and reading EOY test scores in eighth grade and all student survey outcomes in eighth grade (*G8 Survey Sample*). Students who skipped or repeated grades for each subsample are also excluded.

To test whether students who just pass the ACCESS score reclassification threshold systematically differ from those who just fail to pass were similar in terms of inclusion in each analytic sample, I estimate the following model:

$$O_i = f(A_{iGrade3}) + \delta C_{iGrade3} + \boldsymbol{X_i'\theta} + \varepsilon_i \qquad (4)$$

In this model, the outcome variable O_i represents students' inclusion in one of the four analytic samples. I include the same covariates as those included in the model represented by equation (1) except that the only covariates in vector $\boldsymbol{X}_i'$ are just: dummy variables for the interaction between students' minimum-score ACCESS domain and the school year he or she took the 3rd-grade ACCESS exam; and linear terms for each student's scale scores on the four primary and the overall composite ACCESS domains. I display the results from estimating these models in Table 1.

[Insert Table 1 about here.]

From the results in Table 1, for three of the four samples, I conclude that students on either side of the reclassification threshold look similar to one another. Of the four optimal bandwidth models that I estimated to test for equivalence of students in terms of inclusion into analytic samples, all signaled equivalence except the model for the G8 noncognitive sample. Most sensitivity checks (i.e., using a local quadratic polynomial instead of a local linear polynomial; using a uniform kernel; varying bandwidths)—in particular for the G5 and G8 achievement samples—report similar results. In Online Appendix Table 4, I show largely comparable results when estimating equation (4) for outcomes that capture each individual aspect contributing to exclusion from the analytic samples (i.e., Does the student have baseline demographic data? Did the student leave the district? Did the student have achievement and/or noncognitive outcome data?). Notably, results from Online Appendix Table 4 suggest the primary driving difference between ELs included and excluded from the G8 noncognitive sample is that those excluded lack data on noncognitive outcomes. I consider the implications of this particular finding for my main results in the Discussion section.

[Insert Table 2 about here.]

Using the model represented by equation (4), I also test for equivalence, within sample, on different student demographic characteristics, with O_i now capturing students' gender, identification as Asian non-Hispanic or Hispanic (the two largest EL populations in WCPSS), or disability status. As seen in Table 2, for the G8 achievement and noncognitive samples, ELs barely passing the ACCESS reclassification threshold are more likely to be Hispanic; for the G5 noncognitive sample, those barely passing are less likely to have a disability. Thus, to increase precision of estimates and to account for these differences I include baseline covariates in my models, as noted above. In totality, however, results from these models generally yield little evidence that ELs who just pass reclassification thresholds in third grade differ systematically on observables from those who just fail to. Furthermore, the primary concern for the internal validity of my RD estimates is that non-equivalence on baseline characteristics between those just passing and those just failing to pass the EL reclassification threshold signals that other unobservable characteristics or treatments besides reclassification are occurring that impact students' outcomes. Yet the process of how the ACCESS is scored again suggests that manipulation of other factors at the reclassification threshold is unlikely.

Finally, in Table 3 below I provide the descriptive characteristics of students whose 3rd-grade ACCESS score place them either below or above the reclassification threshold. Two details are worth highlighting. First, these two groups of students demonstrate observable differences on average across each analysis subsample, highlighting the importance of using the RD approach for arriving at unbiased estimates of the impact of reclassification on outcomes for the most similar students on either side of the ACCESS reclassification thresholds—those just above and those just below. Second, because of when WCPSS began collecting student survey data, I do not have short- and long-term noncognitive outcomes for every cohort of 3rd-grade ELs who take the ACCESS exam. Notably, the G8 noncognitive sample contains cohorts of students from the 2009–2010 through 2011–2012 school years, and the G5 noncognitive sample contains cohort of students from the 2012–2013 through 2014–2015 school years. As such, my study is not a true longitudinal investigation of impacts of reclassification on noncognitive outcomes (more temporal overlap is found for the achievement samples).

[Insert Table 3 about here.]

Results

Achievement Outcomes

In Table 4, I present estimates of the impact of passing the ACCESS score threshold for reclassification on students' achievement outcomes in fifth and eighth grade (see equation [1]). Each column represents variations of the same model with the outcomes predicted listed on the leftmost column. Column 1 shows impact estimates from my preferred model, which is estimated using: a sample of students determined by a data-driven bandwidth-identification process (Calonico et al., 2014), a local linear relationship between students' ACCESS scores and outcomes, a triangular kernel, and covariates to improve precision. The bandwidth and resulting student sample size for these models are also reported. Across columns, I explore the sensitivity of impact estimates to adjustments to this preferred model by: varying the bandwidth; removing the triangular kernel weight; excluding covariates; and considering a local quadratic relationship between the ACCESS score and outcomes.

[Insert Table 4 about here.]

As depicted in Column 1 of the table, I find that students who just pass the reclassification threshold in third grade score significantly higher in fifth grade on the reading EOY standardized test in WCPSS. Specifically, students who just pass the threshold score nearly .1-*SD*s higher on the reading EOY; notably, though the exact magnitude fluctuates slightly, this impact estimate is consistently positive and generally significant across sensitivity models. For 5th-grade mathematics EOY, this impact of just passing the reclassification threshold in third grade is also positive, but slightly attenuated (and thus insignificant in my preferred specification) relative to the effect for reading EOY scores. Finally, for the G8 achievement sample, passing the threshold again predicts higher achievement—and significantly—in both subjects. For reading and mathematics, the impact of reclassification is approximately .07 SDs and .14 SDs respectively.

In Table 5, I provide estimates of the average impact of each year of reclassification after third grade on students' EOY mathematics and reading standardized test scores in fifth and eighth grade. These effects come from estimation of the 2SLS RD-IV model represented by equations (2) and (3) above using data from the G5 and G8 achievement samples; I report results from both stages using my preferred model, mirroring the specifications for models captured by Column 1 in Table 3 (i.e., local linear; optimal bandwidth; triangular kernel; and covariates included).

[Insert Table 5 about here.]

The first stage estimates in Table 5 reflect models with students' EL classification status following the 3rd-grade ACCESS as an outcome (see equation [2])—formally testing the relationships seen in Figure 3. I find that students who just pass the reclassification threshold in third grade are reclassified through fifth grade for about 1.3 years more than students who do not and are reclassified through eighth grade for about 1.6 years more than students who do not. These differences reinforce the use of an RD-IV approach to specifically assess the average impact of years of reclassification (as opposed to merely passing all reclassification thresholds in third grade) on student outcomes.

From the second-stage estimates displayed in the table, I show that each year of being reclassified as fully English proficient has substantial positive consequences for ELs' achievement outcomes. For mathematics and reading EOY outcomes measured in fifth grade, the average per-year impact of reclassification (specifically in upper elementary grades) is approximately .04 and .08 SDs, respectively. Students who are not reclassified in third grade are predicted to differ by up to .08 and .16 SDs, on average, in reading and mathematics achievement if they remain classified as ELs for the following two years (i.e., fourth and fifth grade). Using the eighth grade EOY second stage per-year reclassification estimates, I find that students who are not reclassified in third grade are predicted to differ by up to .20 and .45 SDs, on average, if they remain classified as ELs for the following five years (i.e., fourth through eighth grade).

Noncognitive Outcomes

In Tables 6 and 7, I present results from estimation of equation (1) using students' noncognitive outcomes as my outcome variables. Again, my preferred model is depicted in Column 1, but results from sensitivity checks are shown across columns. First, I estimate the impact of just passing the ACCESS score reclassification threshold on students' achievement outcomes, but with the new, restricted sample of students in the G5 and G8 survey samples. For G5, those just passing the threshold continue to demonstrate higher achievement outcomes than those that do not; for G8 this is true for the mathematics EOY, but not for the reading EOY. Overall, however, estimates are generally insignificant, attenuated, and/or less precise due to considerably smaller sample sizes.

[Insert Table 6 about here.]

[Insert Table 7 about here.]

For noncognitive outcomes, I find that just passing the reclassification threshold in WCPSS significantly improves 5th-grade students' out-of-school engagement but decreases the level of challenge for the work their teachers assign. These impact estimates are large ($|\delta| > .08$-*SDs*) and in general fairly robust to model changes (see Columns 2 through 8 in Table 6). For in-school engagement and grit, impacts are slightly smaller and insignificant across models. Finally, I find suggestive negative effects of passing the reclassification threshold for peer relationships and suggestive positive effects for family support, but impacts are not significant for my preferred model (Column 1); across models, these point estimates are again large ($|\delta| > .08$-*SDs*) and occasionally significant, but less precisely estimated than those for out-of-school engagement and challenging work.

For the cohort of ELs that I have 8th-grade noncognitive outcomes for, however, many effects appear to "fade out" or even reverse. The observed negative impact of passing the reclassification threshold on 5th-grade challenging work is attenuated towards zero in eighth grade; the significant positive impact on 5th-grade out-of-school engagement is significant and negative for eighth grade. Overall, the RD-IV second-stage estimates (see Table 8), which describe the average impact of each year of reclassification in upper elementary (or in upper elementary and middle school) grades specifically on noncognitive outcomes, mirror the relationships captured in Tables 6 and 7 (i.e., the ITT estimates): I find significant short-term positive impacts for each year of reclassification on out-of-school engagement and negative impacts for challenging work, with these impacts shrinking or reversing (respectively) for the same noncognitive outcomes measured in eighth grade.

Additional Sensitivity Analyses

In addition to testing the sensitivity of impact estimates to different modeling decisions in Tables 3, 5, and 6, I also explore the sensitivity of my findings when accounting for changes to the ACCESS examination over time. In the 2011–2012 school year, the WIDA Consortium adopted new English language development standards. Furthermore, in the 2016–2017 school year, WIDA also changed scoring of ACCESS so that ELs would have to demonstrate stronger English skills to achieve similar proficiency level scores as before. Though the latter change to the ACCESS does not impact students in my sample (i.e., I do not have outcome data from WCPSS for the 3rd-grade ELs for the 2016-2017 school year onwards), the former could potentially influence findings. As such, I estimate equation (1) for the achievement samples (only ELs that took the ACCESS in third grade after the 2011–2012 school year comprise the G5 noncognitive sample, and splitting the G8 noncognitive sample would result in very small sample sizes), but separately for cohorts of ELs who took the 3rd-grade ACCESS either before 2011–2012 or 2012–2013 onwards. Results from these two groups of students largely communicate the same story for achievement outcomes (see Online Appendix Table 5).

Discussion

The impact of reclassification on achievement is quite large. Depending on the model, I estimate that, on average, each additional year of being reclassified in upper elementary (or in upper elementary and middle school) grades translates to approximate gains between .04 and .09 SDs for mathematics or reading EOY scores, respectively (see Table 5). This result most closely aligns to that of Pope (2016), who finds positive effects of reclassification on ELA outcomes that grow over time for ELs reclassified in elementary school. Conversely, my findings are somewhat at odds with Robinson (2011) and Robinson-Cimpian and Thompson (2016), who find null effects for reclassification in elementary school. However, these authors focused only on year-after impacts for test scores, whereas I estimate effects over a longer time span. Furthermore, I stress again that it is unlikely that any single impact estimate for reclassification exists across contexts, given the myriad contributors to heterogeneity.

Though making comparisons of impact estimates across studies may not be useful, comparing the effect of reclassification on a variety of outcomes within the same context is. Above I show that just passing the ACCESS reclassification threshold in third grade significantly increases ELs' out-of-school engagement, with suggestive positive impacts for reports of family support two years after reclassification; I also observe a negative short-term impact on access to challenging work, with suggestive negative impacts on perceptions of peer relationships (see Table 6). Though no prior classification study has estimated impacts on these measures, comparing the magnitudes to those for achievement outcomes suggests that reclassification has an equal if not even bigger effect on certain noncognitive outcomes. Overall, however, impacts of reclassification appear to shift in the longer-term, as relationships across the board attenuate or reverse for WCPSS

ELs' outcomes measured five years after reclassification. However, I stress that caution should be taken when interpreting the longer-term noncognitive results. As noted earlier, I find some evidence that ELs may be differentially (but necessarily) excluded from my G8 noncognitive sample for not having noncognitive data (see Table 1 and Online Appendix Table 4)—if these excluded students had these data and subsequently been included in models, results could have obviously changed.

Conclusion and Policy Implications

In my analyses, I show that passing the ACCESS threshold for being reclassified as fully English proficient at the end of third grade leads to significant short-term gains in reading standardized test scores and longer-term gains in reading and mathematics standardized test scores in WCPSS. Because I employ an RD approach, my study is one of few studies to arrive at credible causal impacts of reclassification on students' achievement outcomes. Prior research on this topic using RDs argue that identifying non-null impacts of reclassification suggests a better policy regime exists (Robinson, 2011). For example, positive effects on test scores might indicate that students just meeting the reclassification thresholds are more than ready to succeed in school without the instructional supports afforded to ELs. As such, policymakers might reduce the stringency of thresholds to allow more ELs to be reclassified—at least, up until the point where reclassification no longer had a positive effect. This would reduce the burden on schools and teachers to provide the additional instructional supports to a larger population of students.

However, I also highlight Robinson's (2011) argument that null effects of reclassification using an RD approach do not necessarily indicate that the current policy is the best policy. Specifically, reclassifying students as fully English proficient may lead to unintended consequences (Umansky, 2018) not captured by impact evaluations on just achievement outcomes. Reclassification might influence students' noncognitive outcomes or other indicators of future success. For example, I find evidence that each year of reclassification on average improves students' out-of-school engagement in the short-term. Focusing on just positive achievement impacts may thus underestimate the benefits of a policy shift towards less stringent reclassification thresholds. Contrastingly, I find that each year of reclassification on average negatively impacts students' reports on the level of challenge for the work teachers assign them in school. Given this finding, loosening criteria for reclassification may not be without negative consequence. Relatedly, I find that effects attenuate or even reverse when using noncognitive outcomes measured in the longer-term, i.e., those measured five years after reclassification. This result underscores that evaluations should (when possible) track and evaluate outcomes longitudinally as impacts may change as students progress through school—which would again inform a different policy response.

As such, I argue that more complete evaluations of educational agencies' policies should assess the effects of reclassification on a variety of student outcomes—especially when ultimate outcomes are unavailable to investigate. Observing null effects across several intermediate outcomes would lend support to the appropriateness of existing reclassification policies. Alternatively, even one observed non-null effect on an important indicator would suggest that the current policy regime could be improved on. This might mean changing reclassification thresholds, as described above, or addressing the settings that ELs enter and exit after being reclassified. It is worth noting, however, that policy changes may not affect ELs' performance similarly across outcomes. Furthermore, in some cases, policymakers may need to weigh impact estimates for certain outcomes more heavily than others to arrive at a decision; in these instances, follow-up evaluations should follow Robinson-Cimpian and Thompson's study (2016) and investigate how policy shifts change the impact of reclassification.

As this extension in considering multiple types of outcomes has clear implications for the educational experiences of ELs, future work should build on my analyses—the first to attempt to gather evidence on the impact of reclassification on students' in-school engagement, out-of-school engagement, grit, familial support, or perceptions of their peers and schoolwork. This might entail, for example, a qualitative study that first aims to assess whether students' survey responses truthfully match their experiences (i.e., and are not influenced by different self-report biases) and, if so, to then develop a better understanding of why reclassified students in WCPSS report less challenging work being assigned by their teachers and worse relationships with peers in order to determine how the district might better support them. Indeed, results from this study converge with anecdotal evidence from the district that reclassified ELs may be benefiting academically from the increased inclusion in mainstream instructional settings but may be experiencing worse relations with their teachers and peers due to fewer touch points with dedicated EL classrooms. With reclassification policies being determined by the state and not the district, WCPSS might change the instructional settings available to EL and reclassified students leveraging results using the student survey outcomes—results that emerged only after considering multiple measures.

Relatedly, my results do not necessarily challenge prior research arguing that ELs may be experiencing bias and discrimination from others in school (Umansky, 2016, 2018). Reclassified students in WCPSS may in fact be experiencing fewer negative interactions with teachers and non-EL students, but the loss of access to special programming dedicated specifically for ELs (Harklau, 1994; Suárez-Orozoco, Suárez-Orazoco, & Todorova, 2008) may dominate any positive labeling effects tied to status gain (Link & Phelan, 2001). The exact mechanisms again should be teased out more clearly with future research.

As is true in all studies using regression discontinuity approaches, the generalizability of my estimates is limited. Specifically, my estimates of the impact of reclassification on students' achievement outcomes apply only to students who just pass all reclassification thresholds in WCPSS and those who just fail at least one. Furthermore, because I employ a regression discontinuity instrumental variables approach, the average per-year estimates specifically apply only to students who "comply" with their treatment. However, students whose English proficiency place them just above or below the reclassification cutoff are those whose educational experiences policymakers are most able to change, i.e., by changing the reclassification criteria. As such, the estimates I show here are only limited by the fact that they do not pertain to ELs who are far from demonstrating English proficiency or who are extremely fluent. Relatedly, these estimates should not necessarily be considered as capturing the impact of losing intensive EL instruction. As noted in my background section, in WCPSS, teachers place each EL's instructional needs into three categories of intensity. ELs just below the reclassification threshold are likely those receiving the least intense instructional supports.

As the population of English learners in the US grows, more rigorous studies assessing the impacts of the policies influencing ELs' educational instruction are needed. Furthermore, these impact evaluations need to consider a wide range of outcome variables tracked over time in order to ensure that changes to policies are not misinformed.

References

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), 427-445.

Blanchard, S., & Muller, C. (2015). Gatekeepers of the American Dream: How teachers' perceptions shape the academic outcomes of immigrant and language-minority students. *Social Science Research, 51,* 262–275.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal, 14*(4), 909–946.

Carlson, D., & Knowles, J. E. (2016). The effect of English Language Learner reclassification on student ACT scores, high school graduation, and postsecondary enrollment: Regression discontinuity evidence from Wisconsin. *Journal of Policy Analysis and Management, 35*(3), 559–586.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics, 126*(4), 1593-1660.

Chin, A. (2015). *Impact of bilingual education on student achievement.* Bonn, Germany: IZAWorld of Labor.

Cimpian, J., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English learner reclassification policy effects across districts. *American Educational Research Journal, 54*(1_suppl), 255S–278S.

Dabach, D. B. (2014). I am not a shelter!: Stigma and social boundaries in teachers' accounts of students' experience in separate "sheltered" English learner classrooms. *Journal of Education for Students Placed at Risk, 19*(2), 98–124.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087.

Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In M.D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs: Theory and Applications* (pp. 281–315). Bingley, United Kingdom: Emerald Publishing Limited.

Galston, W. A. (2007). Civic knowledge, civic education, and civic engagement: A summary of recent research. *International Journal of Public Administration, 30*(6-7), 623–642.

Harklau, L. (1994). Jumping tracks: How language-minority students negotiate evaluations of ability. *Anthropology & Education Quarterly, 25*(3), 347–363.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation* (No. w7867). Cambridge, MA: National Bureau of Economic Research.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Huang, H. (2018). *How do students feel about their schools? WCPSS student survey results: 2017-18.* Wake County Public School System: Cary, NC.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.

Jacob, R., Zhu, P., Somers, M. A., & Bloom, H. (2012). *A practical guide to regression discontinuity.* New York, NY: MDRC.

Johnson, A. (2019). The impact of English learner reclassification on high school reading and academic progress. *Educational Evaluation and Policy Analysis.*

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151.

Kanno, Y., & Kangas, S. E. N. (2014). I'm not going to be, like, for the AP: English language learners' limited access to advanced college-preparatory courses in high school. *American Educational Research Journal, 51*(5), 848–878.

Lau v. Nichols, 414 U.S. 563 (1974).

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics, 142*(2), 655–674.

Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27*(1), 363-385.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698–714.

Musu-Gillette, L., de Brey, C., McFarland, J., Hussar, W., Sonnenberg, W., & Wilkinson-Flicker, Sidney. (2017). *Status and trends in the education of racial and ethnic groups 2017.* NCES 2017-051. Washington, DC: National Center for Education Statistics.

Poulou, M. S. (2017). Social and emotional learning and teacher–student relationships: Preschool teachers' and students' perceptions. *Early Childhood Education Journal, 45*(3), 427-435.

Pope, N. G. (2016). The marginal effect of K-12 English language development programs: Evidence from Los Angeles Schools. *Economics of Education Review, 53*, 311–328.

Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, *5*(1), 83-104.

Reyes, M., & Hwang, N. (2019). Middle School Language Classification Effects on High School Achievement and Behavioral Outcomes. *Educational Policy.*

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis, 33*(3), 267–292.

Robinson-Cimpian, J. P., & Thompson, K. D. (2016). The effects of changing test-based policies for reclassifying English learners: Reclassification policy effects. *Journal of Policy Analysis and Management, 35*(2), 279–305.

Shin, N. (2018). The effects of the initial English Language Learner classification on students' later academic outcomes. *Educational Evaluation and Policy Analysis, 40*(2), 175–195.

Suárez-Orozco, C., Suárez-Orozco, M. M., & Todorova, I. (2010). *Learning a new land: Immigrant students in American society.* Cambridge, MA: Belknap Press.

Thompson, K. D. (2015). Questioning the long-term English learner label: How categorization can blind us to students' abilities. *Teachers College Record, 117*(12), n12.

Umansky, I. M. (2016). To be or not to be EL: An examination of the impact of classifying students as English learners. *Educational Evaluation and Policy Analysis, 38*(4), 714–737.

Umansky, I. M. (2018). According to plan? Examining the intended and unintended treatment effects of EL classification in early elementary and the transition to middle school. *Journal of Research on Educational Effectiveness, 11*(4), 588-621.

Umansky, I. M., & Dumont, H. (2019). *English learner labeling: How English learner status shapes teacher perceptions of student skills and the moderating role of bilingual instructional settings.* (EdWorkingPaper No. 19-94). Providence, RI: Annenberg Brown University.

U.S. Department of Education. (2016). *Non-regulatory guidance: English learners and Title III of the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act*. Washington, DC: Authors.

U.S. Department of Education. (2012). *National evaluation of Title III implementation—Report on state and local implementation.* Washington, DC: Policy and Program Studies Service, Office of Planning, Evaluation, and Policy Development, U.S. Department of Education.

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis, 37*(4), 612-637.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis, 38*(1), 148-170.

WIDA. (2019). *Interpretive guide for school reports: Kindergarten-Grade 12.* Madison, WI: Authors.

Yoon, B. (2008). Uninvited guests: The influence of teachers' roles and pedagogies on the positioning of English language learners in the regular classroom. *American Educational Research Journal,* 45(2), 495-522.

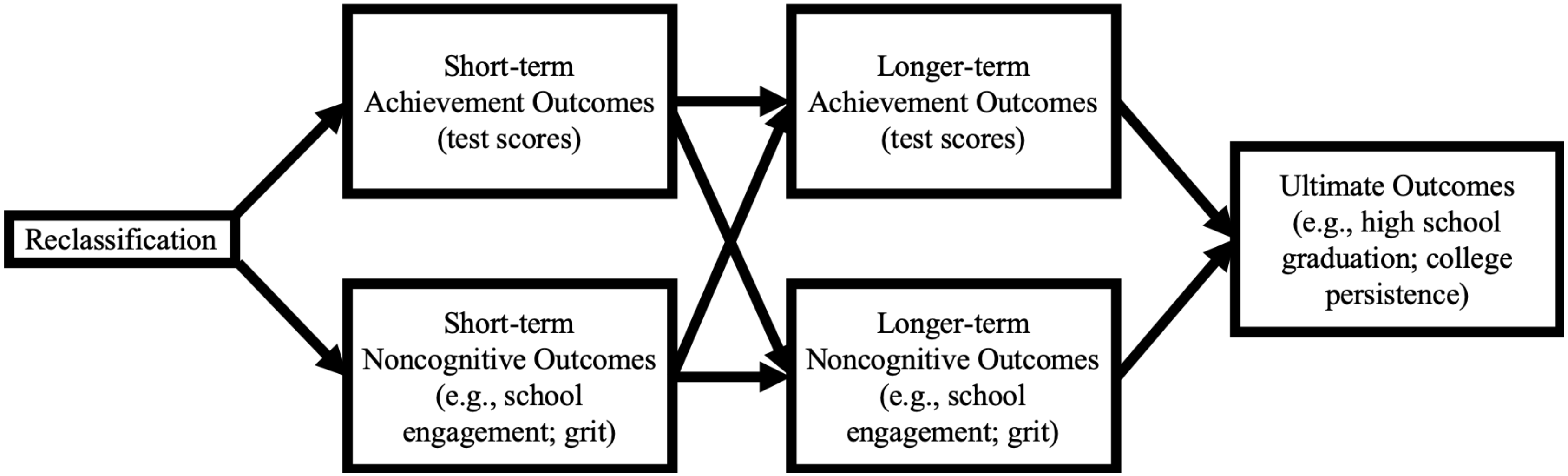
Reclassification
Short-term Achievement Outcomes (test scores)
Short-term Noncognitive Outcomes (e.g., school engagement; grit)
Longer-term Achievement Outcomes (test scores)
Longer-term Noncognitive Outcomes (e.g., school engagement; grit)
Ultimate Outcomes (e.g., high school graduation; college persistence)

Figure 1.

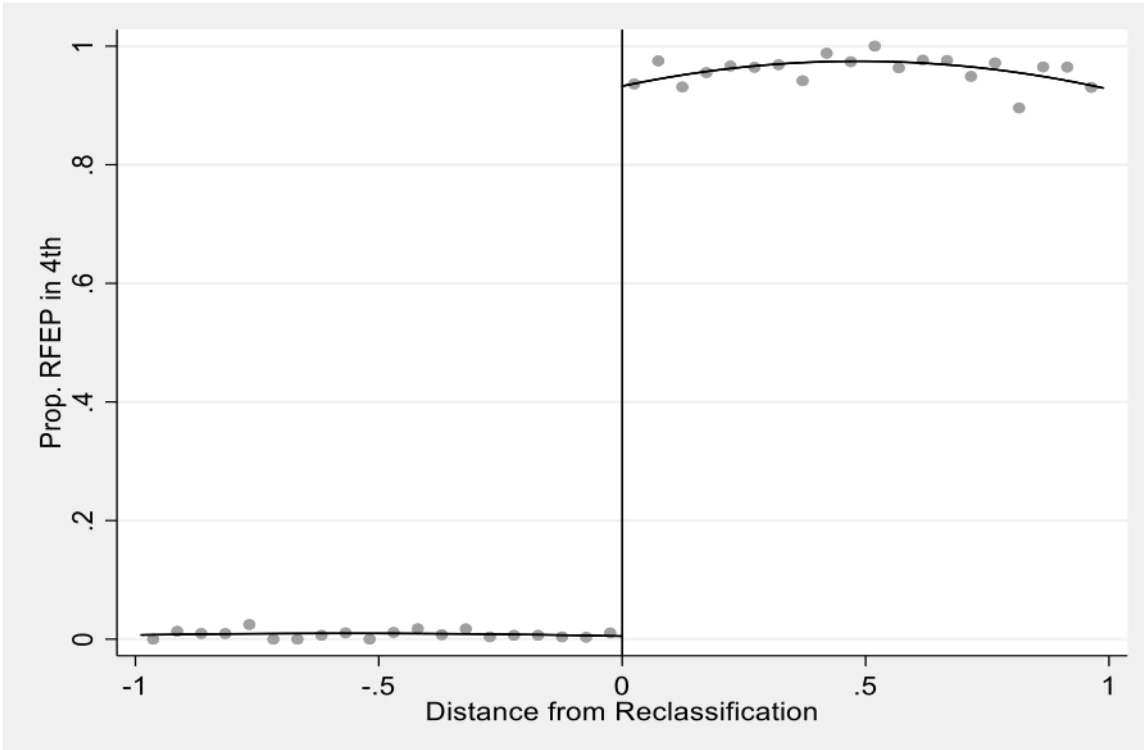
Prop. RFEP in 4th
0
.2
.4
.6
.8
1
-1
-.5
0
.5
1
Distance from Reclassification

Figure 2.

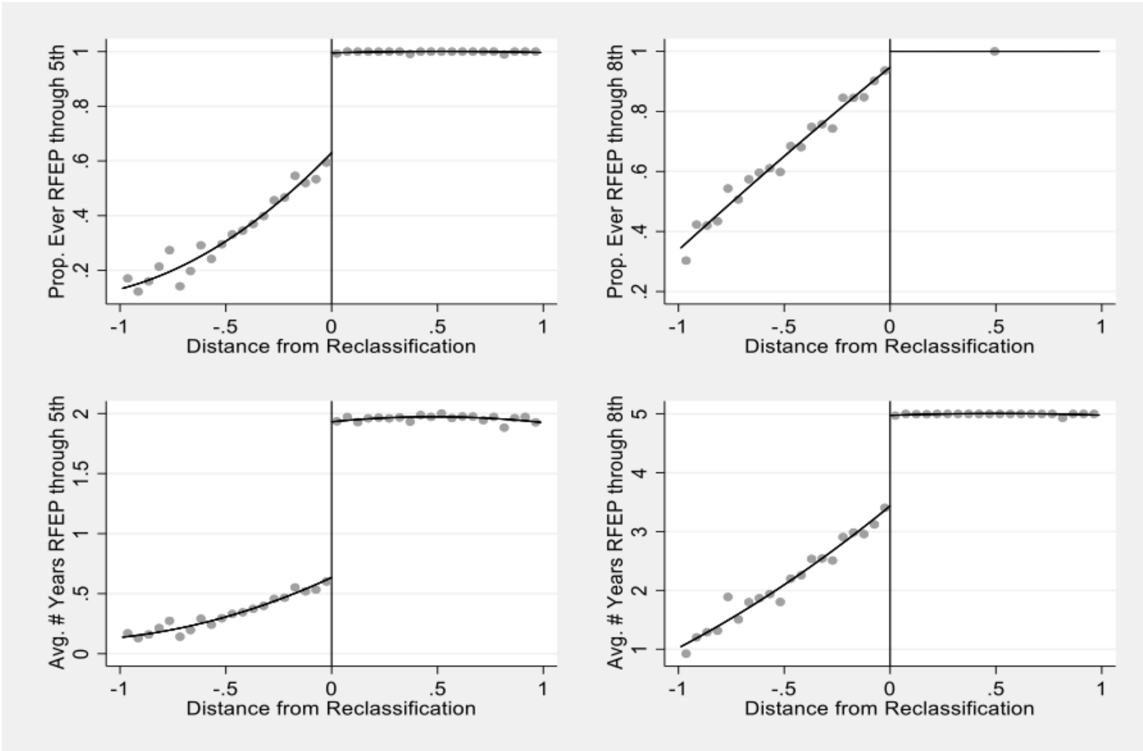
Prop. Ever RFEP through 5th
Prop. Ever RFEP through 8th
Avg. # Years RFEP through 5th
Avg. # Years RFEP through 8th
Distance from Reclassification
Distance from Reclassification
Distance from Reclassification
Distance from Reclassification

Figure 3.

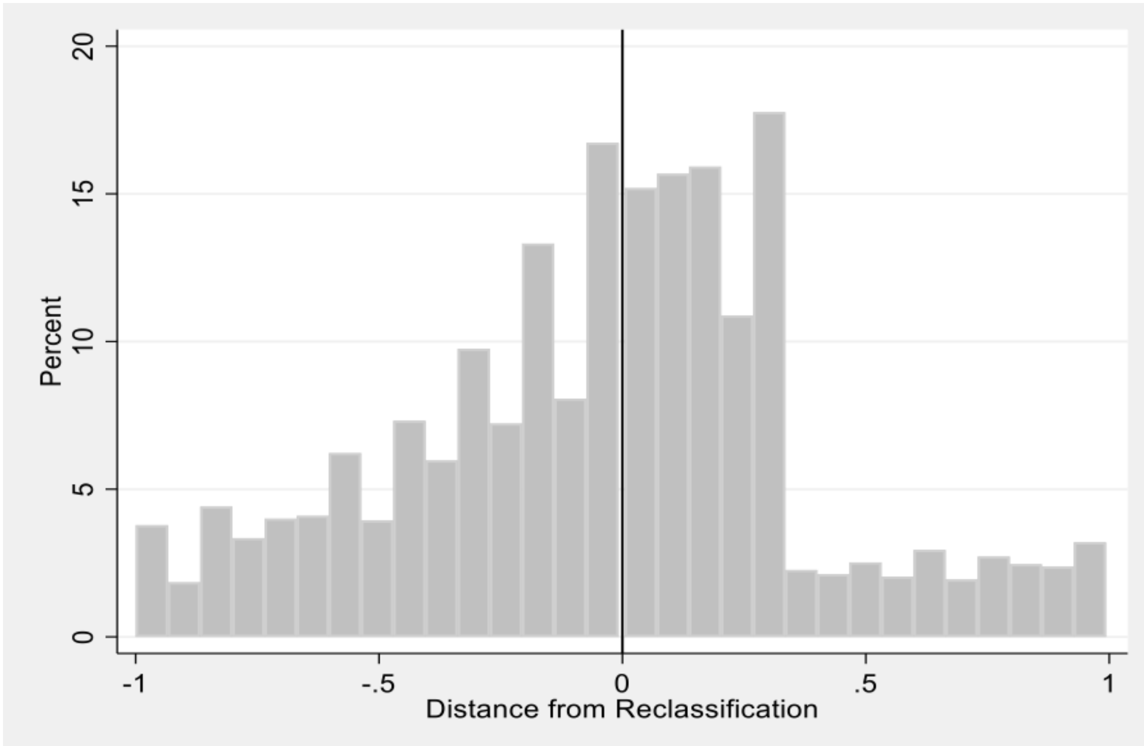

Figure 4.

Figure Captions

Figure 1. Schema of relationships that link reclassification to student outcomes over time.

Figure 2. Binned scatterplot of EL students' reclassification status in 4th-grade against their 3rd-grade ACCESS score. Students with scores at or above zero, marked by vertical line, should be reclassified as fully English proficient. RFEP = "Reclassified as Fully English Proficient".

Figure 3. Binned scatterplots of EL students' reclassification status in 5th- and 8th-grade against their 3rd-grade ACCESS score. Students with scores at or above zero, marked by vertical line, should be reclassified as fully English proficient. RFEP = "Reclassified as Fully English Proficient". Only one bin of observations is observed for the plot of RFEP by eighth grade against distance from reclassification because all students with ACCESS scores above zero are reclassified by eighth grade.

Figure 4. Histogram of EL students' 3rd-grade ACCESS score. Students with scores at or above zero, marked by a vertical black line, should be reclassified as fully English proficient.

Table 1. Effect of scoring above reclassification threshold from intent-to-treat models predicting students' sample inclusion.

	1	BW	n	2	3	4	5	6	7
In G5 Achievement Sample	-0.00767	0.430	5921	-0.000879	-0.00165	-0.00913	0.00574	0.0153	-0.00156
	(0.0158)			(0.0176)	(0.0171)	(0.0152)	(0.0190)	(0.0226)	(0.0181)
In G8 Achievement Sample	0.000634	0.368	3730	-0.00463	-0.00239	-0.00170	0.00194	-0.00692	0.00746
	(0.0311)			(0.0290)	(0.0329)	(0.0292)	(0.0402)	(0.0450)	(0.0365)
In G5 Noncog Sample	-0.00017	0.604	2648	0.0220	0.00742	-0.00241	0.0992**	0.154***	0.0692~
	(0.0355)			(0.0479)	(0.0372)	(0.0335)	(0.0375)	(0.0408)	(0.0390)
In G8 Noncog Sample	0.0417~	0.272	1964	0.0341	0.0567**	0.0351~	0.0335	0.0495	0.0335
	(0.0231)			(0.0283)	(0.0213)	(0.0205)	(0.0319)	(0.0363)	(0.0262)
Local polynomial	1			1	1	1	2	2	2
Bandwidth	Opt			Opt	Opt*.08	Opt*1.2	Opt	Opt*.08	Opt*1.2
Kernel	Tri			Uni	Tri	Tri	Tri	Tri	Tri

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Local polynomial: 1 = linear, 2 = quadratic. Bandwidth: Opt = optimal. Kernel: tri = triangular, Uni = uniform. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 2. Effect of scoring above reclassification threshold from intent-to-treat models predicting students' baseline characteristics.

	1	BW	n	2	3	4	5	6	7
				Panel A. G5 Achievement Sample					
Male	-0.00960	0.513	5704	0.00111	-0.0124	-0.00206	-0.0287	-0.0255	-0.0172
	(0.0202)			(0.0230)	(0.0203)	(0.0199)	(0.0235)	(0.0259)	(0.0234)
Asian non-Hispanic	-0.0177	0.343	4640	-0.0263	-0.0106	-0.0145	-0.0138	-0.00590	-0.0152
	(0.0164)			(0.0186)	(0.0147)	(0.0176)	(0.0191)	(0.0205)	(0.0191)
Hispanic	0.0216	0.300	3713	0.0160	0.0115	0.0322	-0.00445	-0.0228	0.0125
	(0.0141)			(0.0164)	(0.0117)	(0.0203)	(0.0237)	(0.0243)	(0.0243)
Has disablity	0.000980	0.688	6689	-0.0145	-0.00220	0.00627	-0.0148	-0.0154	-0.00578
	(0.0152)			(0.0164)	(0.0164)	(0.0147)	(0.0176)	(0.0181)	(0.0186)
				Panel B. G8 Achievement Sample					
Male	-0.0188	0.618	3639	-0.0221	-0.0256	-0.0188	-0.0377	-0.0460	-0.0304
	(0.0329)			(0.0363)	(0.0337)	(0.0318)	(0.0362)	(0.0369)	(0.0372)
Asian non-Hispanic	-0.0217	0.481	3232	-0.0278	-0.0271	-0.0180	-0.0325	-0.0341	-0.0278
	(0.0237)			(0.0260)	(0.0242)	(0.0233)	(0.0277)	(0.0294)	(0.0267)
Hispanic	0.0447*	0.368	2872	0.0447*	0.0474*	0.0426*	0.0530*	0.0488*	0.0486*
	(0.0195)			(0.0216)	(0.0202)	(0.0204)	(0.0226)	(0.0227)	(0.0229)
Has disablity	0.00195	0.458	3208	0.0140	-0.00257	0.00506	-0.00540	-0.0137	-0.00172
	(0.0160)			(0.0230)	(0.0153)	(0.0163)	(0.0168)	(0.0165)	(0.0181)
				Panel C. G5 Noncog Sample					
Male	-0.0442	0.416	1730	-0.0541	-0.0457	-0.0415	-0.0680	-0.0929	-0.0671
	(0.0421)			(0.0442)	(0.0462)	(0.0384)	(0.0563)	(0.0687)	(0.0564)
Asian non-Hispanic	-0.0216	0.397	1672	-0.0337	-0.0167	-0.0164	-0.00575	0.00617	-0.0202
	(0.0253)			(0.0271)	(0.0215)	(0.0256)	(0.0249)	(0.0235)	(0.0285)
Hispanic	0.0283	0.326	1552	0.00934	-0.00676	0.0378	-0.0476	-0.104**	-0.00874
	(0.0323)			(0.0354)	(0.0250)	(0.0380)	(0.0319)	(0.0381)	(0.0364)
Has disablity	-0.0652*	0.413	1730	-0.0533	-0.0746**	-0.0551~	-0.101***	-0.101**	-0.0939**
	(0.0303)			(0.0353)	(0.0283)	(0.0318)	(0.0301)	(0.0333)	(0.0306)
				Panel D. G8 Noncog Sample					
Male	-0.0352	0.567	1906	-0.0417	-0.0369	-0.0294	-0.0453	-0.0337	-0.0350

	(0.0502)			(0.0522)	(0.0524)	(0.0488)	(0.0576)	(0.0563)	(0.0591)
Asian non-Hispanic	-0.0213	0.502	1820	-0.0415	-0.0320	-0.0145	-0.0359	-0.0406	-0.0264
	(0.0394)			(0.0463)	(0.0413)	(0.0382)	(0.0464)	(0.0497)	(0.0450)
Hispanic	0.0556~	0.452	1766	0.0762*	0.0665*	0.0400	0.0816*	0.0863*	0.0622~
	(0.0322)			(0.0357)	(0.0321)	(0.0333)	(0.0367)	(0.0354)	(0.0373)
Has disablity	0.0250	0.408	1676	-0.00152	0.00906	0.0282	0.00480	-0.0265	0.0196
	(0.0215)			(0.0199)	(0.0194)	(0.0213)	(0.0226)	(0.0201)	(0.0257)
Local polynomial	1			1	1	1	2	2	2
Bandwidth	Opt			Opt	Opt*.08	Opt*1.2	Opt	Opt*.08	Opt*1.2
Kernel	Tri			Uni	Tri	Tri	Tri	Tri	Tri

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Local polynomial: 1 = linear, 2 = quadratic. Bandwidth: Opt = optimal. Kernel: tri = triangular, Uni = uniform. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 3. Descriptive statistics for 3rd-grade ELs in WCPSS.

	G5 Achievement		G8 Achievement		G5 Noncognitive		G8 Noncognitive	
Year range: 3rd Grade Cohorts	2009	2016	2009	2013	2013	2015	2010	2012
Has ACCESS scores?	0.911		0.912		0.923		0.911	
Has baseline demo data?	0.974		0.966		0.974		0.970	
Did not leave district?	0.908		0.837		0.911		0.845	
Has achievement data?	0.905		0.857		0.939		0.898	
Has noncognitive data?					0.826		0.684	
In sample?	0.772		0.693		0.682		0.546	
	Below	Above	Below	Above	Below	Above	Below	Above
Male	0.567	0.506	0.554	0.496	0.592	0.511	0.54	0.496
Hispanic	0.825	0.694	0.834	0.685	0.846	0.722	0.825	0.667
Asian, non-Hispanic	0.076	0.171	0.072	0.17	0.075	0.163	0.074	0.185
With Disability	0.145	0.043	0.151	0.039	0.208	0.046	0.135	0.038
Math EOY	-0.862	-0.118	-0.676	-0.016	-0.865	-0.115	-0.711	-0.007
	(0.817)	(0.841)	(0.79)	(0.878)	(0.823)	(0.84)	(0.755)	(0.839)
Reading EOY	-1.154	-0.266	-0.911	-0.088	-1.178	-0.271	-0.821	0.007
	(0.815)	(0.777)	(0.8)	(0.804)	(0.824)	(0.777)	(0.812)	(0.783)
Challenging Work					-0.011	-0.052	0.038	0.075
					(1.043)	(1.064)	(0.921)	(0.906)
In-school Engagement					0.387	0.453	-0.451	-0.506
					(0.815)	(0.845)	(0.934)	(0.98)
Grit					0.171	0.391	-0.359	-0.234
					(0.859)	(0.965)	(0.903)	(1.003)
Out-of-school Engagement					0.017	0.239	-0.213	-0.053
					(0.968)	(0.856)	(1.052)	(1.032)
Peer Relationships					0.06	0.182	-0.201	-0.051

	(0.981)	(0.958)	(1.015)	(0.988)
Family Support	0.023	0.258	-0.237	-0.133
	(0.996)	(0.863)	(1.052)	(1.055)

Table 4. Effect of scoring above reclassification threshold from intent-to-treat models predicting students' achievement outcomes after third grade for G5 and G8 achievement samples.

	1	BW	n	2	3	4	5	6	7	8
					Panel A. G5 Achievement Sample					
G5 Reading EOY	0.0996***	0.300	3713	0.0631	0.113***	0.110***	0.0610	0.114*	0.158**	0.0543
	(0.0283)			(0.0404)	(0.0292)	(0.0272)	(0.0399)	(0.0501)	(0.0506)	(0.0552)
G5 Math EOY	0.0503	0.393	4964	0.0151	0.0634	0.0900*	0.0419	0.157**	0.183**	0.114~
	(0.0505)			(0.0624)	(0.0468)	(0.0437)	(0.0477)	(0.0580)	(0.0605)	(0.0620)
					Panel B. G8 Achievement Sample					
G8 Reading EOY	0.0680~	0.369	2872	0.0696~	0.0625	0.102**	0.0584	0.0954*	0.147***	0.0539
	(0.0409)			(0.0398)	(0.0485)	(0.0313)	(0.0418)	(0.0431)	(0.0375)	(0.0534)
G8 Math EOY	0.143**	0.499	3333	0.0846	0.110~	0.152**	0.144**	0.159*	0.186**	0.146*
	(0.0513)			(0.0727)	(0.0605)	(0.0562)	(0.0473)	(0.0630)	(0.0623)	(0.0610)
Local polynomial	1			1	1	1	1	2	2	2
Bandwidth	Opt			Opt	Opt	Opt*.08	Opt*1.2	Opt	Opt*.08	Opt*1.2
Kernel	Tri			Tri	Uni	Tri	Tri	Tri	Tri	Tri
Covariates	Yes			No	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Local polynomial: 1 = linear, 2 = quadratic. Bandwidth: Opt = optimal. Kernel: tri = triangular, Uni = uniform. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 5. Coefficients and standard errors from two-stage least square RD-IV models using years reclassified to predict achievement outcomes after third grade for G5 and G8 achievement samples.

	Reading EOY	Math EOY
	Panel A. G5 Achievement Sample	
2SLS - 1st stage	1.321***	1.317***
	(0.0303)	(0.0262)
2SLS - 2nd stage	0.0754***	0.0382
	(0.0213)	(0.0388)
Bandwidth	0.300	0.393
N	3713	4964
	Panel B. G8 Achievement Sample	
2SLS - 1st stage	1.579***	1.589***
	(0.0809)	(0.0689)
2SLS - 2nd stage	0.0431	0.0897**
	(0.0262)	(0.0346)
Bandwidth	0.369	0.499
N	2872	3333

Note: 1st stage coefficients reflect difference in years reclassified between students just scoring above the reclassification threshold and those scoring just below the threshold. 2nd stage coefficients reflect the impact of each year of reclassification on outcomes. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Model specifications include: local polynomial: linear; bandwidth: optimal; kernel: triangular; covariates: included. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 6. Effect of scoring above reclassification threshold from intent-to-treat models predicting students' achievement and noncognitive outcomes after third grade for G5 noncognitive sample.

	1	BW	n	2	3	4	5	6	7	8
Reading EOY	0.0563	0.350	1552	0.0632	0.0523	0.0641	0.0702	0.110	0.146	0.0987
	(0.0827)			(0.0489)	(0.0969)	(0.0574)	(0.0862)	(0.141)	(0.135)	(0.134)
Math EOY	0.101	0.613	1966	0.0207	0.110	0.0998	0.101	0.111	0.0951	0.107
	(0.0805)			(0.0617)	(0.0750)	(0.0956)	(0.0689)	(0.169)	(0.174)	(0.159)
Challenging Work	-0.119***	0.407	1730	-0.150***	-0.0989*	-0.100**	-0.133***	-0.129**	-0.110*	-0.129**
	(0.0289)			(0.0351)	(0.0432)	(0.0344)	(0.0265)	(0.0481)	(0.0534)	(0.0409)
In-school Engagement	-0.0503	0.577	1911	-0.0347	-0.0683	-0.0526	-0.0480	-0.0432	-0.00424	-0.0659
	(0.0897)			(0.0893)	(0.0879)	(0.0972)	(0.0813)	(0.114)	(0.120)	(0.110)
Grit	0.0471	0.342	1552	0.00643	-0.0482	0.0998~	0.0111	-0.00462	0.0972	-0.0312
	(0.0540)			(0.0583)	(0.0643)	(0.0570)	(0.0546)	(0.0788)	(0.0724)	(0.0681)
Out-of-school Engagement	0.0826~	0.370	1643	0.119*	0.113*	0.127**	0.0717~	0.119*	0.175***	0.0929~
	(0.0437)			(0.0476)	(0.0577)	(0.0424)	(0.0418)	(0.0512)	(0.0440)	(0.0521)
Peer Relationships	-0.132	0.460	1789	-0.144~	-0.159*	-0.117	-0.144~	-0.115	-0.0900	-0.161~
	(0.0840)			(0.0751)	(0.0770)	(0.0904)	(0.0800)	(0.0988)	(0.0969)	(0.0959)
Family Support	0.0840	0.365	1643	0.0954	0.101	0.149*	0.0445	0.149~	0.227**	0.0859
	(0.0669)			(0.0651)	(0.0698)	(0.0650)	(0.0654)	(0.0801)	(0.0829)	(0.0812)
Local polynomial	1			1	1	1	1	2	2	2
Bandwidth	Opt			Opt	Opt	Opt*.08	Opt*1.2	Opt	Opt*.08	Opt*1.2
Kernel	Tri			Tri	Uni	Tri	Tri	Tri	Tri	Tri
Covariates	Yes			No	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Local polynomial: 1 = linear, 2 = quadratic. Bandwidth: Opt = optimal. Kernel: tri = triangular, Uni = uniform. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 7. Effect of scoring above reclassification threshold from intent-to-treat models predicting students' achievement and noncognitive outcomes after third grade for G8 noncognitive sample.

	1	BW	n	2	3	4	5	6	7	8
Reading EOY	-0.0281	0.273	1247	-0.0551	0.00912	-0.0503	-0.0395	-0.0757	-0.0523	-0.111
	(0.0569)			(0.0437)	(0.0759)	(0.0617)	(0.0560)	(0.0691)	(0.0697)	(0.0696)
Math EOY	0.0696	0.274	1247	0.00318	0.0590	0.0804	0.0662	0.0532	0.0454	0.0292
	(0.0460)			(0.0585)	(0.0528)	(0.0502)	(0.0449)	(0.0609)	(0.0530)	(0.0669)
Challenging Work	-0.0270	0.431	1698	-0.0203	-0.0694	-0.0279	-0.0225	-0.0300	-0.0347	-0.00831
	(0.0520)			(0.0583)	(0.0552)	(0.0586)	(0.0497)	(0.0704)	(0.0681)	(0.0662)
In-school Engagement	-0.0558	0.465	1766	-0.0466	-0.0147	-0.0293	-0.0650	-0.0512	-0.0415	-0.0572
	(0.0827)			(0.0822)	(0.0725)	(0.0828)	(0.0779)	(0.0974)	(0.105)	(0.0968)
Grit	-0.0325	0.454	1766	-0.0281	-0.00293	-0.0222	-0.0324	0.0117	0.0743	-0.00638
	(0.0713)			(0.0633)	(0.0892)	(0.0815)	(0.0641)	(0.104)	(0.106)	(0.0942)
Out-of-school Engagement	-0.132*	0.497	1820	-0.114	-0.128	-0.126~	-0.120~	-0.101	-0.0896	-0.123~
	(0.0665)			(0.0729)	(0.0789)	(0.0647)	(0.0676)	(0.0809)	(0.107)	(0.0675)
Peer Relationships	-0.0642	0.426	1676	-0.0641	-0.0392	-0.0524	-0.0501	-0.0512	0.0380	-0.0754
	(0.0905)			(0.103)	(0.106)	(0.107)	(0.0784)	(0.141)	(0.135)	(0.130)
Family Support	-0.0682	0.338	1507	-0.0750	-0.0969	-0.0778	-0.0818	-0.0615	-0.0939	-0.0547
	(0.0683)			(0.0717)	(0.0836)	(0.0817)	(0.0690)	(0.0926)	(0.108)	(0.0829)
Local polynomial	1			1	1	1	1	2	2	2
Bandwidth	Opt			Opt	Opt	Opt*.08	Opt*1.2	Opt	Opt*.08	Opt*1.2
Kernel	Tri			Tri	Uni	Tri	Tri	Tri	Tri	Tri
Covariates	Yes			No	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Local polynomial: 1 = linear, 2 = quadratic. Bandwidth: Opt = optimal. Kernel: tri = triangular, Uni = uniform. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 8. Coefficients and standard errors from two-stage least square RD-IV models using years reclassified to predict achievement and noncognitive outcomes after third grade for G5 and G8 noncognitive samples.

	Reading EOY	Math EOY	Challenging Work	In-school Engagement	Grit	Out-of-school Engagement	Peer Relationships	Family Support
				Panel A. G5 Achievement Sample				
2SLS - 1st stage	1.222***	1.257***	1.230***	1.253***	1.222***	1.224***	1.242***	1.223***
	(0.0404)	(0.0420)	(0.0421)	(0.0425)	(0.0401)	(0.0411)	(0.0433)	(0.0409)
2SLS - 2nd stage	0.0461	0.0807	-0.0964***	-0.0402	0.0385	0.0675~	-0.106	0.0687
	(0.0687)	(0.0645)	(0.0243)	(0.0714)	(0.0446)	(0.0370)	(0.0654)	(0.0568)
Bandwidth	0.350	0.613	0.407	0.577	0.342	0.370	0.460	0.365
N	1552	1966	1730	1911	1552	1643	1789	1643
				Panel B. G8 Achievement Sample				
2SLS - 1st stage	1.537***	1.535***	1.458***	1.458***	1.458***	1.459***	1.459***	1.489***
	(0.112)	(0.111)	(0.0826)	(0.0778)	(0.0792)	(0.0755)	(0.0834)	(0.104)
2SLS - 2nd stage	-0.0183	0.0453	-0.0185	-0.0383	-0.0223	-0.0903~	-0.0440	-0.0458
	(0.0388)	(0.0300)	(0.0361)	(0.0565)	(0.0494)	(0.0473)	(0.0613)	(0.0479)
Bandwidth	0.273	0.274	0.431	0.465	0.454	0.497	0.426	0.338
N	1247	1247	1698	1766	1766	1820	1676	1507

Note: 1st stage coefficients reflect difference in years reclassified between students just scoring above the reclassification threshold and those scoring just below the threshold. 2nd stage coefficients reflect the impact of each year of reclassification on outcomes. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Model specifications include: local polynomial: linear; bandwidth: optimal; kernel: triangular; covariates: included. *~p<.1; *p<.05; **p<.01; **p<.001*

Online Appendix: Tables

Table 1. Survey text and reliabilities of composite outcome variables.

Item	Text
In-school-engagement composite, Cronbach's Alpha=.89	
tsr_treatfair	Overall, adults at my school treat students fairly.
tsr_listen	Adults at my school listen to the students.
tsr_care	At my school, teachers care about students.
tsr_need	My teachers are there for me when I need them.
tsr_rulesfair	The school rules are fair.
tsr_honestopen	Overall, my teachers are honest and open with me.
tsr_enjoytalk	I enjoy talking to the teachers here.
tsr_safe	I feel safe at school.
tsr_interestme	Most teachers are interested in me as a person, not just as a student.
csrw_testsgood	The tests in my classes do a good job of measuring what I am able to do.
csrw_learnimportant	Most of what is important to know you can learn in school.
csrw_gradesaccurate	The grades in my classes do a good job of measuring what I am able to do.
csrw_learnfuture	What I am learning in my classes will be important in my future.
csrw_learnfun	Learning is fun because I get better at something.
csrw_stuvoice	I feel like I have a say about what happens to me at school.
Grit composite, Cronbach's Alpha=.75	
dgs_distract	New ideas and projects sometimes distract me from previous ones (reversed).
dgs_dontdiscourage	Setbacks (delays and obstacles) do not discourage me. I bounce back from disappointments faster than most people.
dgs_obsessed	I have been obssessed with a certain idea or project for a short time but later lost interest (reversed).
dgs_hardworker	I am a hard worker.

dgs_changegoals	I often set a goal but later choose to pursue (follow) a different one (reversed)
dgs_diffprojectfocus	I have difficulty maintaining (keeping) my focus on projects that take more than a few months to complete (reversed).
dgs_finishbeing	I finish whatever I begin.
dgs_dilligent	I am diligent (hard working and careful).
csrw_checkwork	After finish my schoolwork, I check it over to see if it is correct.
csrw_checkunderstand	When I do schoolwork, I check to see whether I understand what I am doing.
csrw_performhardwork	When I do well in school, it is because I work hard.
rs_workexpectations	I work hard to meet my teachers' expectations.
ce_attnnews	I pay attention to what is going on in the news.
ce_boringpolitics	I think politics and government are boring.
ce_commprojects	I participate in projects in my community.

Out-of-school engagement composite, Cronbach's Alpha=.81	
fg_edposths	I plan to continue my education following high school.
fg_importantposths	Going to school after high school is important
fg_schfuturegoals	School is important for achieving my future goals.
fg_edopportunities	My education will create many future opportunities for me.
fg_futurehope	I am hopeful about my future.
ce_makedifference	I believe I can make a difference in my community.
ce_vote	When I am old enough, I plan to vote in most elections.
ce_electionpres	I care a great deal about who is elected to be our next president.

Peer relationships composite, Cronbach's Alpha=.84	
psl_stucare	Other students at school care about me.
psl_need	Students at my school are there for me when I need them.
psl_likeme	Other students here like me the way I am.
psl_enjoytalk	I enjoy talking to students here.
psl_respect	Students here respect what I have to say.
psl_friends	I have some friends at school.

Family support composite, Cronbach's Alpha=.87	
fsl_need	My family/guardian(s) are there for me when I need them.
fsl_help	When I have problems at school, my family/guardian(s) are willing to help me.
fsl_goodknow	When something good happens at school, my family/guardian(s) want to know about it.
fsl_keeptry	My family/guardian(s) want me to keep trying when things are tough at school.

Challenge work item	
rs_challengework	My teachers give me challenging work.

Table 2. Factor loadings for five-factor solution from exploratory factor analysis

	In-school Engagement	Grit	Out-of-school Engagement	Peer Relationships	Family Support
tsr_treatfair	.689681	-.0872714	-.070083	.0078347	.0556755
tsr_listen	.6789832	-.0610291	-.0846714	.0100881	.0672036
tsr_care	.7408673	-.1017036	-.0248771	-.0170532	.0700008
tsr_need	.6955433	-.0894468	-.0556389	.0203875	.1042956
tsr_rulesfair	.6667506	.0179268	-.0678555	-.0621132	.0022879
tsr_honestopen	.6459672	-.0707652	-.0424156	.0405132	.0362613
tsr_enjoytalk	.6288818	.007179	-.0108484	.0509012	.0289091
tsr_safe	.5372048	-.0390635	-.0119688	.1566766	.0352049
tsr_interestme	.5833598	-.0225014	.0066398	.1004967	-.016141
csrw_testsgood	.4899209	.0864101	.0391248	.0120727	-.0464251
csrw_learnimportant	.6066378	.0143152	.1165644	-.0503118	-.0634177
csrw_gradesaccurate	.4302225	.157755	.0584806	.0141478	-.0545454
csrw_learnfuture	.5603857	.0214713	.252568	-.0756959	-.0740236
csrw_checkwork	.2605907	.4229773	.0388912	-.0036739	-.0650669
csrw_checkunderstand	.2000051	.3490256	.0951557	.0486396	-.0420365
csrw_learnfun	.4606034	.2125297	.1296641	-.006816	-.0726821
csrw_performhardwork	.1923485	.2637914	.2313893	-.0069196	.0401079
csrw_stuvoice	.3470649	.032555	.0546105	.0390338	.0128545
rs_challengework	.0198046	-.0879886	.1232449	.0070645	.0278131
rs_workexpectations	.2147348	.3743717	.1397071	.0043683	-.0080144
psl_stucare	-.0104793	.0160552	-.0733885	.7787453	-.0266164
psl_need	.0074908	.0249373	-.0421391	.7731446	-.0312491

psl_likeme	-.0481923	.0597066	-.0328871	.7393655	.0098489
psl_enjoytalk	.1269291	-.0489077	.0568374	.6002438	.0239239
psl_respect	.1083874	.0446404	-.1101366	.6544202	-.0121786
psl_friends	-.0335717	-.0622157	.1942552	.4488163	.0680178
fg_edposths	-.0642848	.0357157	.6415969	-.0098401	.0519448
fg_importantposths	-.0371251	.0511667	.6683037	-.0448163	.0085915
fg_schfuturegoals	.0902229	-.0143504	.726809	-.0865105	-.0192996
fg_edopportunities	.0187819	.0146815	.6945711	-.0345683	-.0020401
fg_futurehope	-.0875528	.1348432	.495116	.0452413	.1097766
ce_makedifference	.008475	.2648256	.2831025	.0852366	.0551969
ce_vote	-.0807088	.1824781	.291535	.0578761	.0091252
ce_electionpres	-.0032129	.0991408	.2413613	.0360177	.0298055
ce_attnnews	.047743	.2423982	.1446922	-.0017873	.0120364
ce_boringpolitics	-.183931	-.2539795	-.0170892	.093184	.0325005
ce_commprojects	.1972832	.2846353	.0822953	.0065652	-.0367545
fsl_need	.0352578	-.0407763	.0349294	.0196415	.7188637
fsl_help	.0634844	.0283994	-.0365848	-.0111602	.7454079
fsl_goodknow	.0732318	.0466468	.0311173	-.0022387	.6320614
fsl_keeptry	.035512	-.0030762	.2191671	-.0646104	.5186517
dgs_distract	-.0228706	-.3478368	.1797148	.0488753	-.030355
dgs_dontdiscourage	-.0931547	.1948876	.0763243	.0927042	.0330129
dgs_obsessed	.0303481	-.3137658	.22309	.0371202	-.0755143
dgs_hardworker	-.0942827	.7008065	.0882058	.0305002	.0066564
dgs_changegoals	.0728553	-.2579027	.2555944	.0185776	-.0742419
dgs_diffprojectfocus	-.0001557	-.4867104	.2386301	.0187694	-.0299536
dgs_finishbeing	-.0814382	.5874651	.0131845	.0089408	.0315116
dgs_dilligent	-.0985724	.7070568	.0806199	.0290401	.0148531

Note: Promax rotation.

Table 3. Fit statistics for five-factor solution from confirmatory factor analysis

Fit statistic	Value
RMSEA	0.047
CFI	0.85
TLI	0.84
SRMR	0.049

Table 4. Effect of scoring above reclassification threshold from intent-to-treat models predicting students' detailed sample inclusion characteristics.

	Grade 5 Achievement	Grade 8 Achievement	Grade 5 Noncognitive	Grade 8 Noncognitive
Has baseline demo data?	0.000454	0.000817	0.0103	-0.00284
	(0.00672)	(0.00851)	(0.00997)	(0.00849)
Did not leave district?	0.00168	0.00677	0.0528~	-0.00265
	(0.0154)	(0.0277)	(0.0291)	(0.0318)
Has achievement data?	0.0124	0.00186	-0.00308	0.0281
	(0.0133)	(0.0294)	(0.0187)	(0.0231)
Has noncognitive data?			-0.00237	0.0317
			(0.0335)	(0.0398)

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. Model specifications include: local polynomial: linear; bandwidth: optimal; kernel: triangular. *~p<.1; *p<.05; **p<.01; **p<.001*

Table 5. Effect of scoring above reclassification threshold from intent-to-treat models predicting student achievement before and after ACCESS standards change.

	Before 2011-2012	BW	n	2012-2013 Onwards	BW	n
G5 Achievement Sample:						
Stacked EOY	0.0587~	0.371	3680	0.0536	0.434	6554
	(0.0350)			(0.0606)		
G8 Achievement Sample:						
Stacked EOY	0.131*	0.426	3676	0.0558	0.549	2670
	(0.0406)			(0.0696)		

Note: Each cell represents a different model estimated using *rdrobust*. Standard errors accounting for clustering at the ACCESS score level are displayed in parentheses. For "stacked" models, each student appears in the data once for each outcome (i.e., Mathematics and Reading achievement). Model specifications include: local polynomial: linear; bandwidth: optimal; kernel: triangular. In these models, fixed effects are also included for each outcome type. *~p<.1; *p<.05; **p<.01; **p<.001.*